

Identification of Key Modules, Hub Genes, and Hub Proteins Associated with Glioblastoma Using Weighted Gene Co-expression Network Analysis and Proteomic Analysis

Nasrin Karimi¹, PhD;  Atieh Tavakoli¹, PhD; Hosna Hesammanesh², MSc; Mohadese Valizadeh^{3,4}, PhD; Hamidreza Kheiri Manjili⁵, PhD; Maryam Yousefi⁶, PhD 

¹Department of Biology, Damghan Branch, Islamic Azad University, Damghan, Iran;

²Department of Biology, Ahar Branch, Islamic Azad University, East Azerbaijan, Iran;

³Department of Tissue Engineering and Applied Cell Sciences, School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran;

⁴Men's Health and Reproductive Health Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran;

⁵Department of Pharmaceutical Nanotechnology, School of Pharmacy, Zanjan University of Medical Sciences, Zanjan, Iran;

⁶Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran

Correspondence:

Maryam Yousefi, PhD;
Science and Research Branch, Shohada Hesarak Blvd., Daneshgah Sq., Sattari Highway,
Postal code: 1483536931, Tehran, Iran
Tel: +98 21 44865179

Email: yousefi.maaryam@gmail.com

Received: 30 September 2025

Revised: 20 February 2026

Accepted: 24 April 2026

What's Known

- Glioblastoma (GBM) is a highly aggressive brain tumor with poor prognosis. Previous studies have applied weighted gene co-expression network analysis (WGCNA) and proteomic analyses separately to explore gene expression and protein networks.
- Several dysregulated genes and pathways have been implicated in GBM progression, but consistent hub genes across transcriptomic and proteomic levels remain unclear.

What's New

- This study integrates WGCNA with proteomic analysis across multiple datasets, identifying *ALDH2*, *ARRB1*, and *TPRG1L* as common hub genes consistently downregulated in GBM.
- It highlights novel gene–protein interactions (e.g., FGG and SUCLG2 linked to *ALDH2/ARRB1*), providing potential biomarkers and therapeutic targets for GBM diagnosis and treatment.

Abstract

Background: Glioblastoma (GBM) is the most aggressive primary brain tumor with a poor prognosis and limited effective therapeutic options. This study aimed to identify key genes and proteins involved in GBM progression and recurrence using integrated weighted gene co-expression network analysis (WGCNA) and proteomic analyses.

Methods: Raw data from nine datasets (GSE100675, GSE111260, GSE90598, GSE58399, GSE50161, GSE4290, GSE108474, GSE134470, and GSE103227) were used for WGCNA of GBM versus control, and the series matrix of GSE62153 was used for WGCNA of Recurrence versus Primary GBM. Transcriptomic data were analyzed in R software (version 4.x) using the limma, sva, and WGCNA packages. Differentially expressed genes (DEGs) were identified using the moderated *t* test with Benjamini–Hochberg False Discovery Rate (FDR) correction (adjusted *P* < 0.05). Common genes between the selected modules and DEGs were identified. Proteomic data (ProteomeXchange PXD035867) were analyzed using MaxQuant and Perseus, and group comparisons were performed using the two-sample Student's *t* test with FDR = 0.05. Enrichment analyses determined shared pathways. All datasets were retrieved from public databases up to 2023, and the study was conducted as a bioinformatics analysis.

Results: Three WGCNA modules (turquoise, blue, and magenta) were identified in the GBM versus control (G/C) comparison, whereas a single module (blue) was identified in the recurrent versus primary GBMs (R/PG) comparison. The common genes between DEGs (G/C), DEGs (R/PG), WGCNA (G/C), and WGCNA (R/PG) (blue module) were aldehyde dehydrogenase 2 (*ALDH2*), arrestin beta 1 (*ARRB1*), and tumor protein p63 regulated 1-like (*TPRG1L*). DEG analysis demonstrated that *ARRB1*, *ALDH2*, and *TPRG1L* genes were downregulated in GBM compared to controls. Proteomic analysis identified three clusters, with fibrinogen gamma chain (FGG) and succinate-CoA ligase GDP-forming beta subunit (SUCLG2) as proteins related to *ALDH2* and *ARRB1*.

Conclusion: Integrated transcriptomic and proteomic analyses identified *ALDH2*, *ARRB1*, and *TPRG1L* as key genes associated with glioblastoma progression, suggesting that these molecules may serve as potential biomarkers or therapeutic targets for GBM.

Please cite this article as: Karimi N, Tavakoli A, Hesammanesh H, Valizadeh M, Kheiri Manjili HR, Yousefi M. Identification of Key Modules, Hub Genes, and Hub Proteins Associated with Glioblastoma Using Weighted Gene Co-expression Network Analysis and Proteomic Analysis. Iran J Med Sci. doi: 10.30476/ijms.2026.108885.4401.

Keywords • Glioblastoma • Gene expression profiling • Proteomics • Biomarkers

Introduction

Glioblastoma (GBM) is the most common malignant primary human brain tumor and the most aggressive tumor among brain tumors.¹ The incidence rate of GBM is 3.2 cases per 100,000 population.² Due to the very poor prognosis of this cancer, most patients die within 1 year of diagnosis.³ Genetic, epigenetic, and microenvironmental factors contribute to the marked heterogeneity of GBM, which represents a major challenge to effective treatment and a key driver of therapeutic failure. of therapeutic failure.⁴ GBMs tend to recur at the initial site of presentation. Recurrence is frequently observed in patients treated postoperatively with radiotherapy or radiotherapy combined with chemotherapy.⁵

Bioinformatics tools have been widely used to study cancer progression mechanisms with the advent of high-throughput technologies such as gene chips and RNA sequencing. Weighted gene co-expression network analysis (WGCNA) is an analytical method for discovering relationships among gene networks across multiple samples. It integrates highly correlated genes into the same module and associates modules with clinical traits.⁶ Thus, WGCNA is very helpful in identifying candidate biomarkers and is a commonly used method in tumor-related research, including GBM.⁷

Proteomics is a powerful approach for analyzing gene expression profiles of thousands of genes simultaneously and gaining insights into protein function and regulation under different conditions.⁸ These technologies are complementary and, in combination, lead to a better understanding of the regulatory events involved in tumorigenesis.

Therefore, the present study aimed to identify common key genes between WGCNA modules and DEGs in GBM versus control and recurrent versus primary GBM, and to integrate these findings with proteomic data to reveal critical pathways involved in GBM progression.

Materials and Methods

All datasets were retrieved from publicly available databases up to 2023, and the study was conducted as a bioinformatics and proteomics analysis. The flowchart of the bioinformatics methods used in this study is presented in figure 1.

Differentially Expressed Genes (DEG) Screening

Raw data from nine datasets (GSE100675, GSE111260, GSE90598, GSE58399, GSE50161, GSE4290, GSE108474, GSE134470, and

GSE103227) and the series matrix of GSE62153 were downloaded from the Gene Expression Omnibus (GEO) database. Detailed information on the GEO datasets and corresponding platforms is presented in table 1.

In the GSE100675 dataset, expression profiles of lncRNAs and mRNAs were obtained from three glioblastoma multiforme (GBM) tissues, three paired non-tumor tissues, and three normal brain tissues. The GSE111260 dataset includes genome-wide expression data of three normal brain and 67 glioma samples. The GSE90598 dataset consisted of 16 fresh-frozen GBM samples, seven healthy brain tissues, and two cell lines.⁹ These datasets were processed using the oligo package.¹⁰

The GSE58399 dataset included 105 GBM cases, while GSE50161 provided gene expression profiles from 130 surgical tumor and normal brain samples.¹¹ In the GSE4290 dataset, gene expression data from 23 non-tumor epilepsy samples and 157 tumor samples (26 astrocytomas, 50 oligodendrogliomas, and 81 GBMs) were available; only GBM and non-tumor samples were included in this study.¹² The GSE108474 dataset comprised 550 samples, including 28 normal samples, 221 GBM, 148 astrocytomas, 67 oligodendrogliomas, and samples with unknown histology.¹³ The GSE134470 dataset contained 58 samples derived from human GBM patients, cell lines, tumor samples, and normal human brain RNA.¹⁴ These datasets were processed and normalized using the affy package with the Robust Multi-array Average (RMA) algorithm.

The GSE103227 dataset, an Agilent single-channel array, provided expression profiles of 26109 mRNAs and 30586 lncRNAs from five GBM samples and five normal brain samples.¹⁵ This dataset was preprocessed using background correction and quantile normalization implemented in the limma R package.¹⁶

Annotation files were downloaded from the GEO database and the Gemma platform. The datasets were then merged and preprocessed to identify differentially expressed genes (DEGs) between GBM and control samples. Batch effects were removed using the ComBat function from the sva package, and genes with low expression intensity were filtered using the genefilter package. DEGs were defined as genes with an adjusted $P < 0.05$ and $|\log_2 \text{fold change} (\log_2 \text{FC})| > 0.2$.

For the analysis of primary versus recurrent GBM, the GSE62153 dataset, including 43 GBM patients (25 primary and 18 recurrent cases), was used.¹⁷ Annotation data were obtained from the illuminaHumanv4.db package.¹⁸

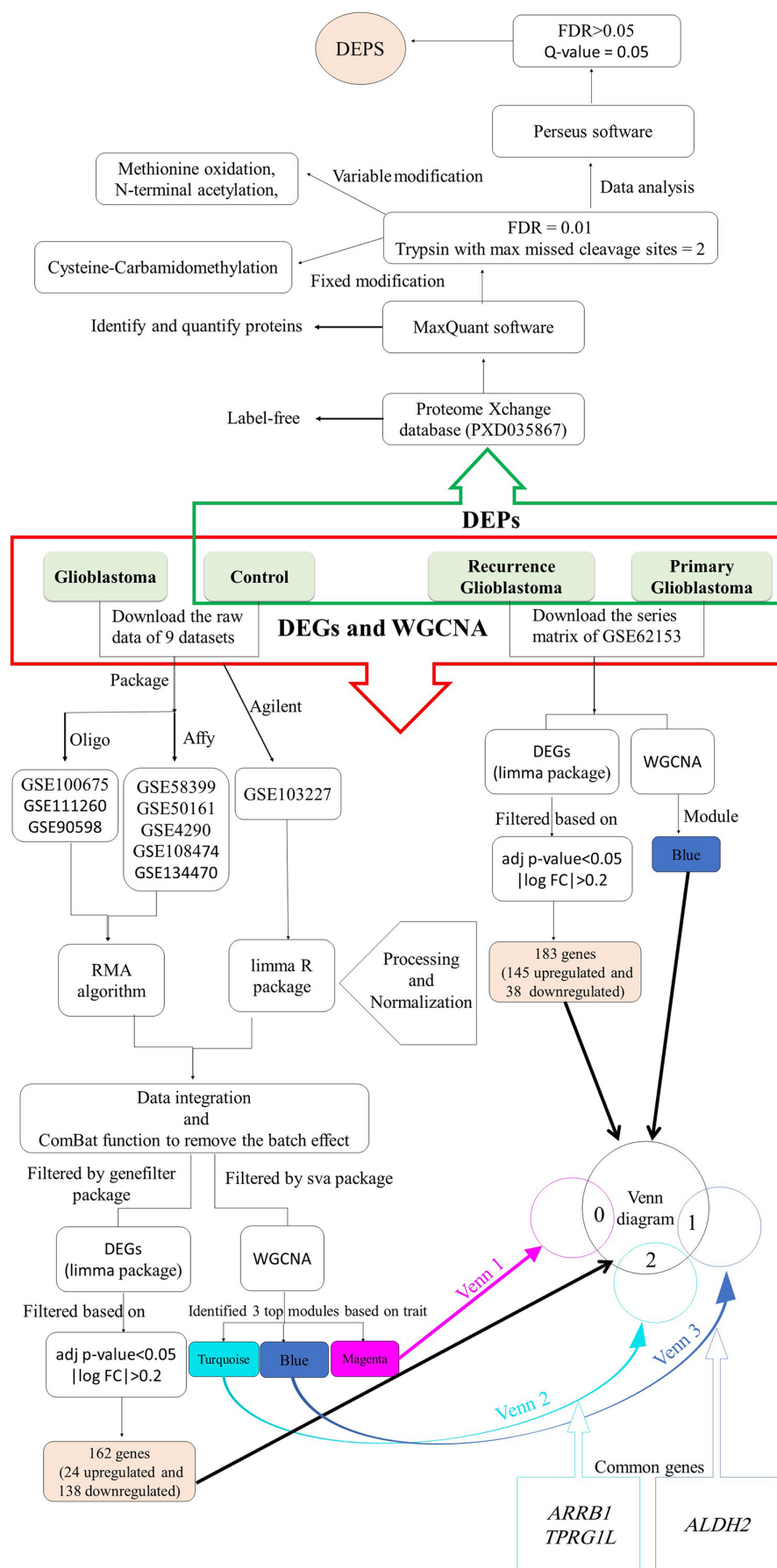


Figure 1: The flowchart illustrates the overall study design and analytical workflow.

Table 1: Information about the Gene Expression Omnibus datasets used for GBM tumor and control samples and for primary and recurrent GBM samples

Datasets	Number of samples			Platforms (Single channel)
	Selected/total	Tumor (GB)	Control	
GSE100675	9/9	3	6	GPL17586 (Affymetrix)
GSE111260	44/70	41	3	GPL5175 (Affymetrix)
GSE90598	23/25	16	7	GPL17692 (Affymetrix)
GSE58399	105/105	105	0	GPL6244 (Affymetrix)
GSE50161	47/130	34	13	GPL570 (Affymetrix)
GSE4290	44/180	77	23	GPL570 (Affymetrix)
GSE108474	249/550	221	28	GPL570 (Affymetrix)
GSE134470	7/58	5	2	GPL6244 (Affymetrix)
GSE103227	10/10	5	5	GPL16956 (Agilent)
Total=594 samples				
Dataset	Recurrence	Primary GBM (PG)		Platforms
GSE62153 (series matrix)	18	25		GPL10558 (Illumina)
Total=43 samples				

Log2 transformation, normalization, and DEG analysis were performed using the limma package (The R Foundation for Statistical Computing, Vienna, Austria),¹⁹ and volcano plots were generated using the EnhancedVolcano package (The R Foundation for Statistical Computing, Vienna, Austria).

WGCNA

WGCNA was performed to identify modules containing genes with the highest co-expression correlations. The aforementioned GSE datasets were used for WGCNA to identify modules associated with primary GBM and recurrent GBM (GSE62153) as well as GBM tumors versus Control samples (GSE100675, GSE111260, GSE90598, GSE58399, GSE50161, GSE4290, GSE108474, GSE134470 and GSE103227). The R WGCNA package was used to perform the analysis of the dataset.²⁰ In this study, we reduced the datasets to relevant genes by selecting probes with a coefficient of variation (CV) greater than 5%. The β parameter, used as a soft thresholding power in WGCNA, strengthens strong correlations while penalizing weak correlations. The Scale Free Topology (SFT) property was used to determine the appropriate β value.²¹ Based on the selected β values, the topological overlap matrix (TOM) and hierarchical clustering dendrograms were constructed to detect gene modules. We restricted module detection to a minimum module size of 30 genes, and modules with high similarity were merged. Finally, hub genes were identified using the chooseTopHubInEachModule function of the WGCNA package.

Common Genes between DEGs and the Modules of WGCNA

The identified DEGs were screened using a threshold of fold change (FC) > 0.2 and 1.5

and an adjusted P value (false discovery rate, FDR) < 0.05. Subsequently, the intersection between DEGs and WGCNA modules related to primary GBM versus recurrent GBM and GBM versus control was performed to identify common genes. The "VennDiagram" web tool was used to identify overlapping genes (<https://bioinformatics.psb.ugent.be/webtools/Venn/>).

To validate our transcriptomic findings, an independent dataset of GBM and normal brain samples was analyzed using the GEPIA2 web tool (<http://gepia2.cancer-pku.cn/>). GEPIA2 provides RNA-seq expression data from TCGA and GTEx projects, allowing comparison of gene expression levels between tumor and normal. Boxplots were generated to visualize the distribution of expression of selected genes, and statistical significance between groups was assessed using the default test implemented in GEPIA2.

Selection of Differentially Expressed Proteins

Protein expression data for GBM were collected from the ProteomeXchange database (PXD035867) and the PRIDE dataset, using label-free quantification (LFQ).⁸ Protein identification and quantification were performed using MaxQuant software with the Andromeda search engine.²² To reduce false-positive identifications, the Target-Decoy strategy was applied, and peptides and proteins were identified at a False Discovery Rate (FDR) of 0.01. For protein digestion, the trypsin enzyme was selected with a maximum of two missed cleavage sites. Methionine oxidation and N-terminal acetylation were set as variable modifications, while cysteine-carbamidomethylation was set as a fixed modification. The "match between runs" option was enabled to reduce missing values across samples.

To assess the quality of the MaxQuant output, the data were further analyzed using Perseus

software.²³ Proteins identified by reverse FASTA sequences, with FDR >0.05, or originating from non-human organisms were excluded. The data were then $\log_2(X)$ -transformed to normalize the intensity values and convert them to a logarithmic scale. Rows were then filtered to remove missing values (NaN), and proteins with fewer than three intensities were excluded from further analysis. Statistical analysis was performed in Perseus using a two-sample test, with a permutation-based false discovery rate (FDR) threshold of 0.05 for identification of differentially expressed proteins (DEPs). Principal component analysis (PCA) and hierarchical clustering were also performed using Perseus. The resulting generic matrix was exported for downstream analysis of the identified DEPs.

Enrichment Analysis

Enrichment Analysis of WGCNA Modules:

After module detection, functional enrichment analysis was conducted to identify the biological processes (BP), molecular function (MF), and cellular components (CC) associated with the genes in each module. The g:Profiler database was used to identify significantly enriched biological categories.²⁴

Enrichment analysis of DEPs: Gene

ontology analysis was performed at the levels of biological process, molecular function, and cellular component to identify pathways associated with the identified DEPs. In addition, the ClueGO plugin in Cytoscape was used for pathway enrichment analysis. Protein-protein interaction and network analyses were performed using the CluePedia Cytoscape plugin (version 5.2.2; INSERM, Paris, France). Edges were filtered using a threshold of 0.6, and interactions of inhibition, binding, action, and post-translational modification (PTM) types were considered. Nodes with the highest degree were selected as hub proteins.

Results

Results of DEGs

We obtained 11902 genes after merging the nine datasets related to GBM patients and controls (GSE100675, GSE111260, GSE90598, GSE58399, GSE50161, GSE4290, GSE108474, GSE134470 and GSE103227). After filtering, 9996 genes remained, of which 162 genes were significantly differentially expressed, including 24 upregulated and 138 downregulated genes. The effectiveness of batch effect removal using the ComBat function is illustrated in figure 2.

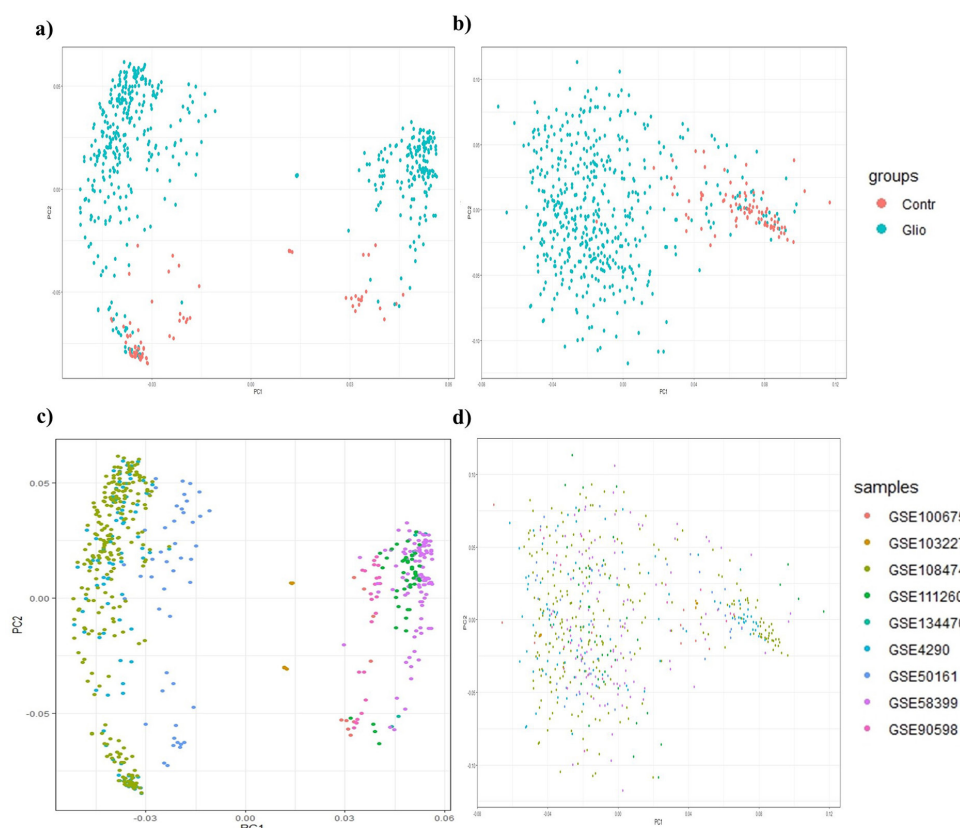


Figure 2: This figure illustrates the effect of the ComBat function on batch correction of the transcriptomic data. (a, c) illustrate the distribution of transcriptomic data before batch effect correction, whereas (b, d) illustrate the distribution after correction using the ComBat function.

In the GSE62153 dataset (primary GBMs vs. recurrent GBMs), a total of 20909 genes were analyzed. After normalization and DEG analysis, 12,946 genes remained, of which 183 genes were identified as significant, including 145 upregulated and 38 downregulated genes. Subsequent analyses revealed that *ARRB1*, *ALDH2*, and *TPRG1L* were common genes between DEGs and WGCNA results.

The DEG analysis demonstrated that *ARRB1*, *ALDH2*, and *TPRG1L* were downregulated in GBM patients compared with controls, whereas their expression levels were upregulated in recurrent GBMs compared with primary GBMs.

As shown in figure 3, volcano plots were constructed for both DEG analyses. The expression levels of *ARRB1*, *ALDH2*, and *TPRG1L* were downregulated in GBM patients compared with controls (figure 3a), while these genes were upregulated in recurrent GBMs compared with primary GBMs (figure 3b). To validate the expression of *ARRB1* and *ALDH2*, an independent dataset including 163 GBM and 207 normal brain samples from GEPIA2 was analyzed (figure 3c). Both *ARRB1* and *ALDH2* were significantly downregulated in GBM compared with normal tissue ($P < 0.05$), confirming our previous transcriptomic findings (figure 3c).

Receiver operating characteristic (ROC) curve analysis of *ARRB1*, *ALDH2*, and *TPRG1L*

in the GBM versus control comparison was performed using the GSE108474 dataset, which had the largest sample size among the nine datasets. The area under the curve (AUC) values for *ARRB1*, *ALDH2*, and *TPRG1L* were 0.92, 0.77, and 0.93, respectively (figures 4a-c). These results suggest that *ARRB1* and *TPRG1L* may serve as potential diagnostic biomarkers for GBM. In addition, the ROC curve of these genes in the primary versus recurrent GBM comparison showed AUC values greater than 0.7, indicating good diagnostic performance (figures 4d-f).

Results of WGCNA

The 11,902 genes obtained from the merged GBM patients versus control datasets and 20,909 genes from the GSE62153 dataset (primary GBMs vs. recurrent GBMs) were independently subjected to WGCNA. Hierarchical cluster analysis of samples and trait heatmaps was performed for each group, and the results are shown in figure 5. To achieve a scale-free topology with $R^2 = 0.9$, a soft-thresholding power (β) of 5 was selected for the GBM versus control datasets (figures 6a and 6b), whereas $\beta = 7$ was selected for the primary GBMs versus recurrent GBMs dataset (figures 6c and 6d). These parameters were used for subsequent network construction. Hierarchical clustering trees were then constructed to identify gene modules.

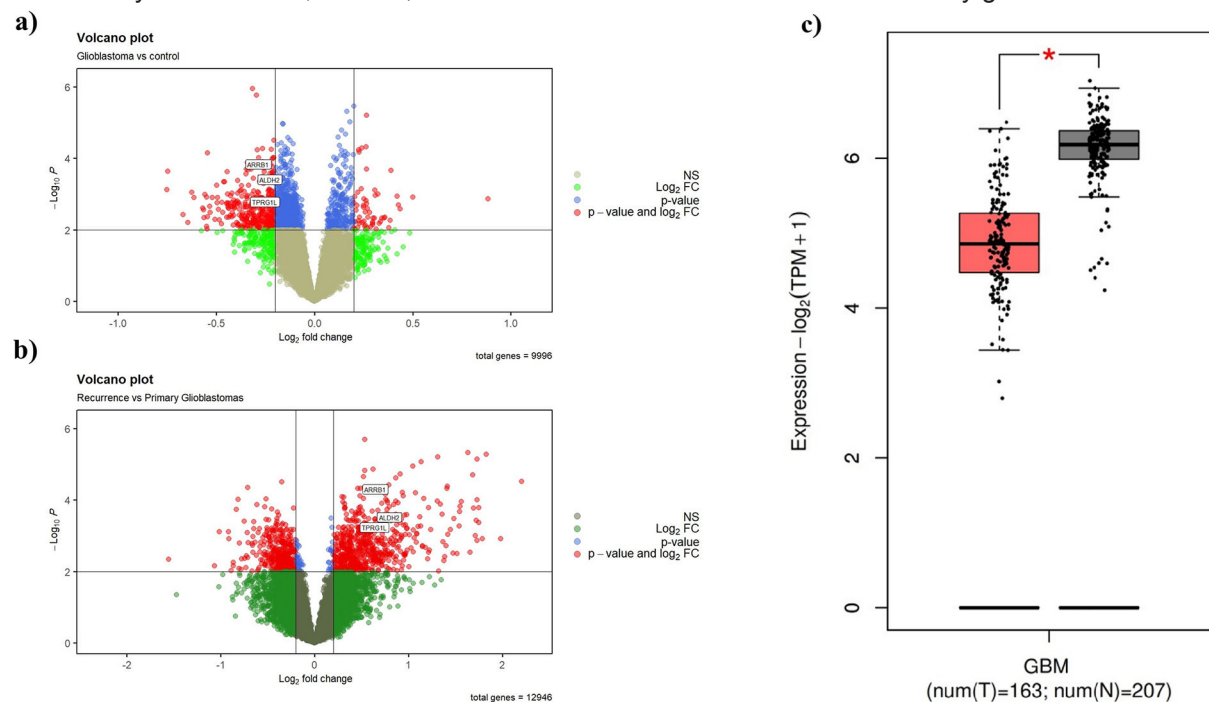


Figure 3: This figure presents volcano plots illustrating differentially expressed genes (DEGs) and the expression patterns of common genes. (a) illustrates the DEGs between GBM and control samples, highlighting the downregulation of *ARRB1*, *ALDH2*, and *TPRG1L* in GBM. (b) illustrates the DEGs between primary and recurrent GBM samples, showing that *ARRB1*, *ALDH2*, and *TPRG1L* are upregulated in recurrent GBM relative to primary GBM. (c) presents a boxplot comparing the expression levels of *ARRB1* and *ALDH2* between GBM (n=163) and normal brain tissue (n=207).

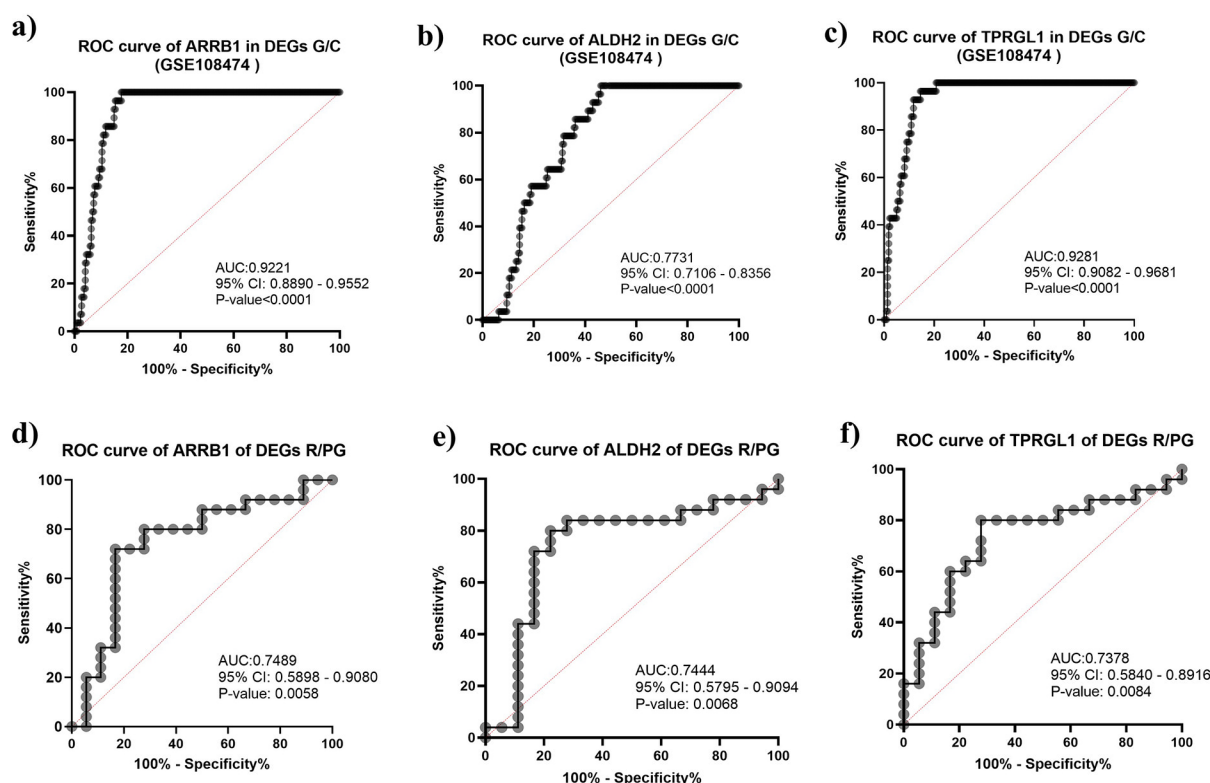


Figure 4: Receiver operating characteristic (ROC) curve analysis of the common genes *ARRB1*, *ALDH2*, and *TPRG1L*. (a–c) illustrate the diagnostic performance of *ARRB1*, *ALDH2*, and *TPRG1L*, respectively, for distinguishing GBM samples from control samples in the GSE108474 dataset. (d–f) illustrate the receiver operating characteristic (ROC) curves of *ARRB1*, *ALDH2*, and *TPRG1L*, respectively, for discriminating recurrent GBM from primary GBM.

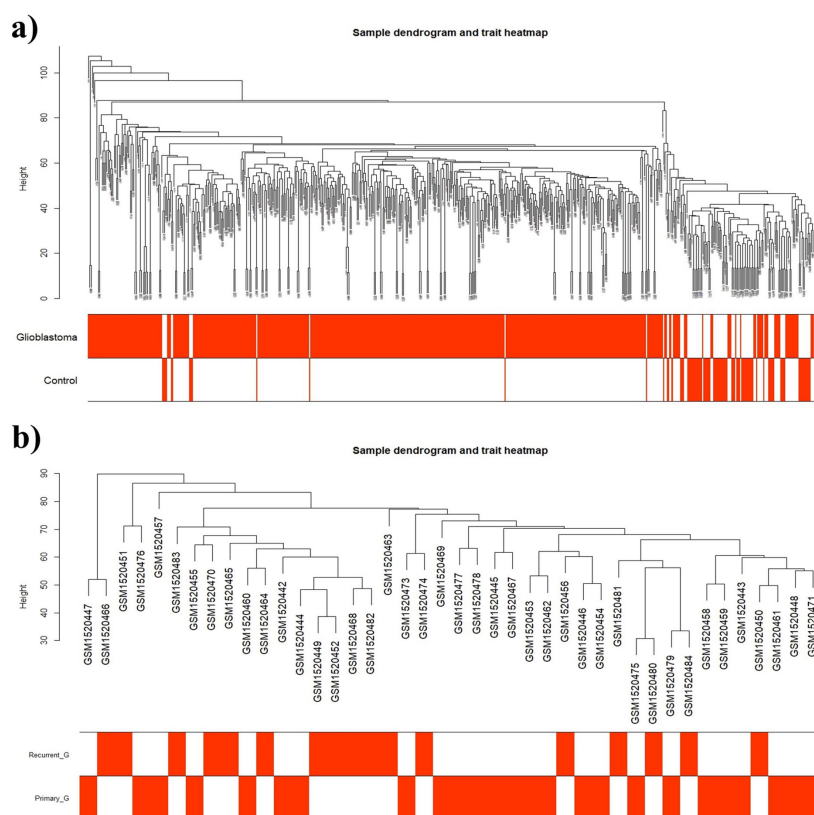
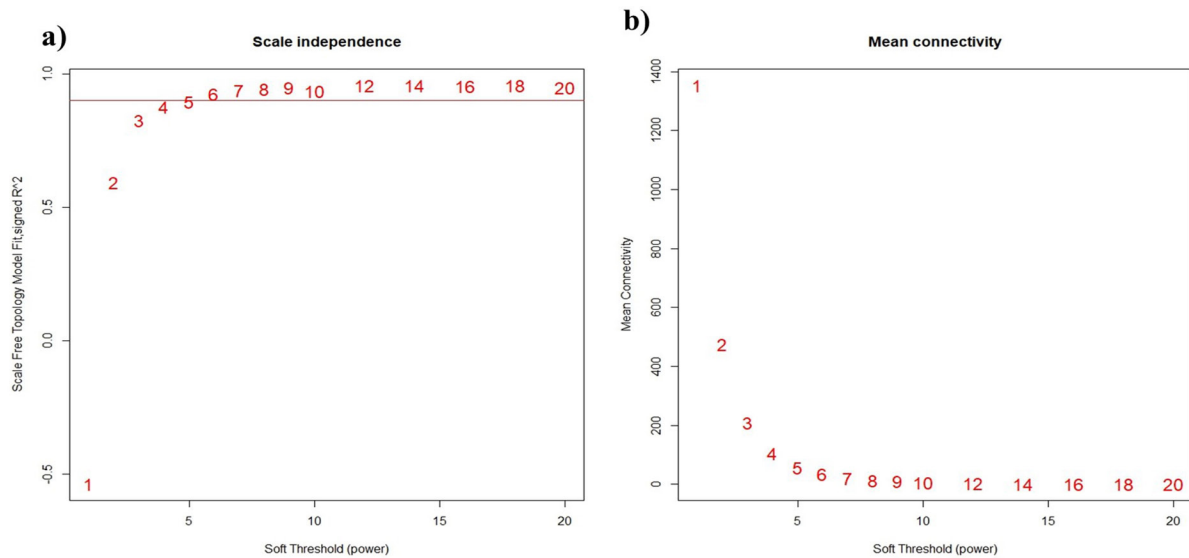


Figure 5: This figure illustrates the hierarchical clustering of samples and the corresponding trait heatmaps for weighted gene co-expression network analysis (WGCNA). The dendrogram leaves represent individual samples, and the color bands denote the associated clinical traits. (a) depicts the clustering of GBM and control samples, whereas (b) depicts the clustering of primary and recurrent GBM samples.

Glioblastoma patients vs. Control



Primary GBMs vs. Recurrent GBMs

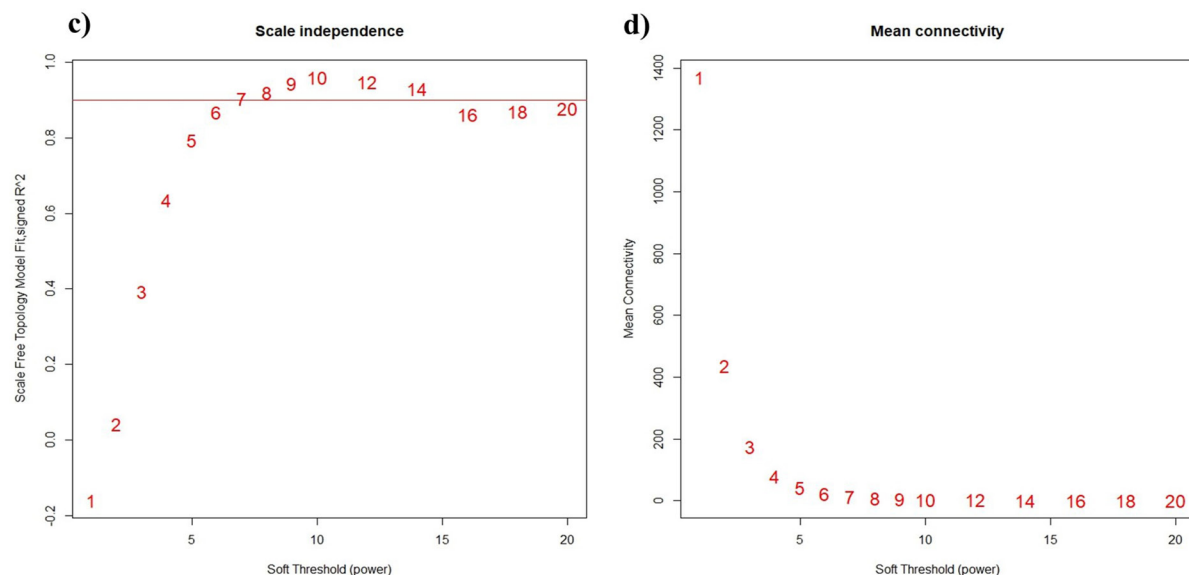


Figure 6: The soft-thresholding power (β) for WGCNA analysis is determined using the scale-free topology criterion ($R^2 = 0.9$). (a) and (b) The analysis for GBM versus control samples, where a power of $\beta=5$ was selected; (c) and (d) the analysis for primary versus recurrent GBM samples, where a power of $\beta=7$ was selected. These β values were used to construct hierarchical clustering trees for module detection in subsequent analyses.

As shown in figure 7, 10 modules were identified in the WGCNA of GBM versus control datasets, while 14 modules were detected in the WGCNA of the GSE62153 dataset (primary GBMs vs. recurrent GBMs). Among these, three modules (turquoise, blue, and magenta) from the GBM versus control analysis and one module (blue) from the primary versus recurrent GBM analysis were selected for further investigation. Module membership is illustrated in figure 8.

The hub genes identified in the WGCNA G/C were *BSN* (turquoise module), *KIF14* (blue module), and *SP110* (magenta module). In

addition, *SEMA4D* was identified as the hub gene of the blue module in the WGCNA R/PG.

Results of Enrichment Analysis of WGCNA Modules

Functional enrichment analysis of the WGCNA G/C modules indicated that genes in the turquoise module were primarily associated with “monoatomic ion transmembrane transporter activity” (molecular function [MF]), “multicellular organism development” (biological process [BP]), and “synapse” (cellular component [CC]), showing the smallest enrichment P value (figure 9a).

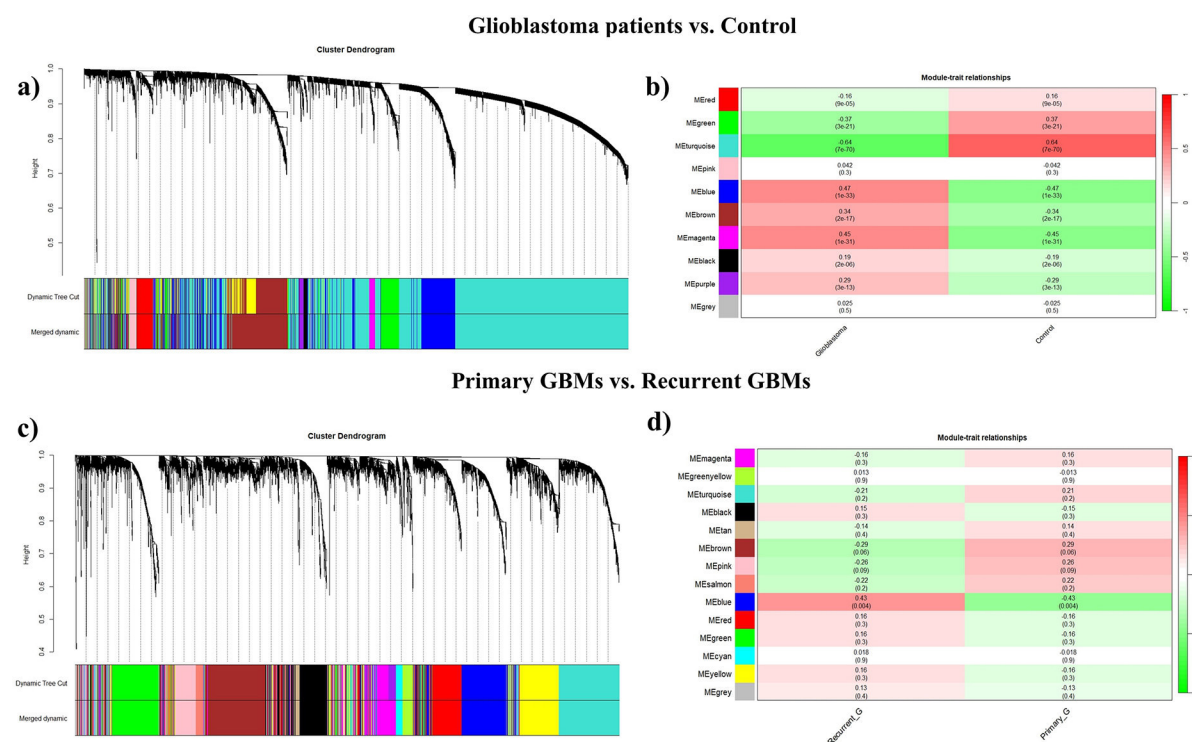


Figure 7: Identification of co-expression modules by WGCNA is illustrated. Ten modules were detected from the analysis of GBM versus control samples using nine merged datasets, and 14 modules were detected from the analysis of primary versus recurrent GBM samples using the GSE62153 dataset.

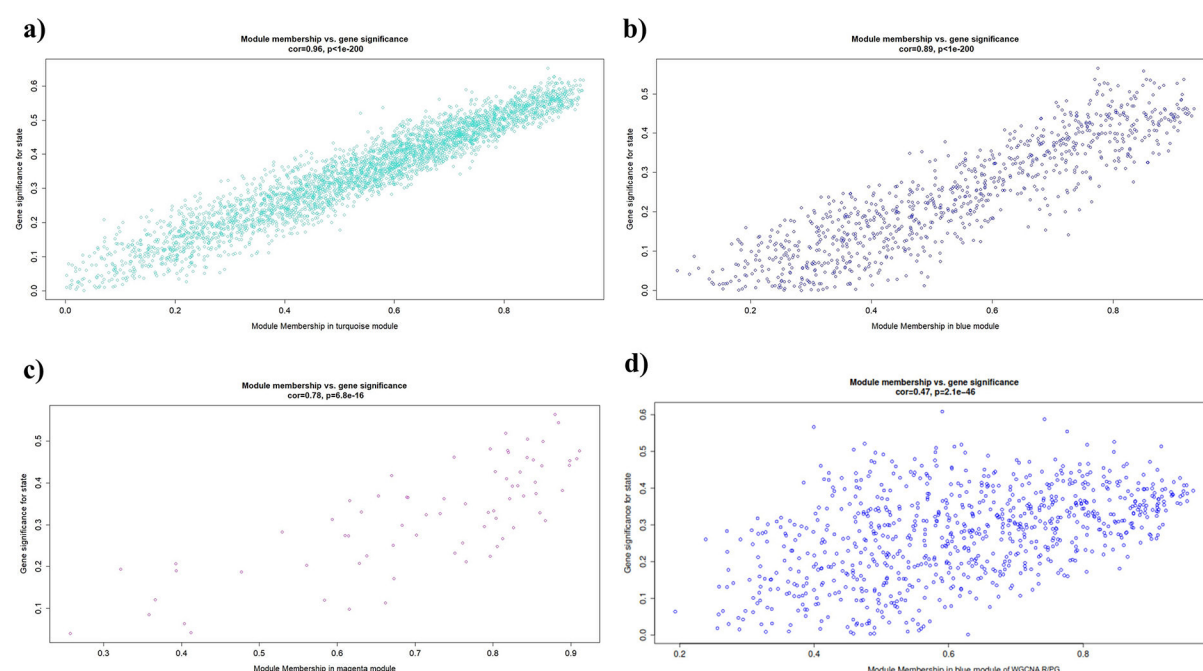
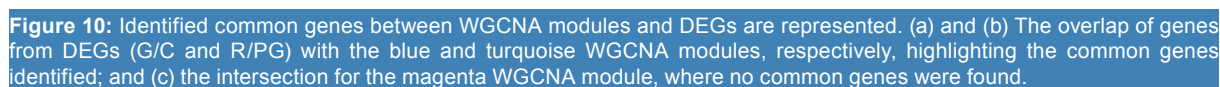
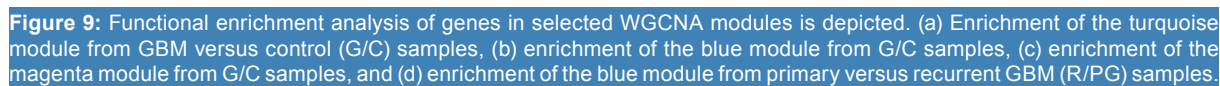


Figure 8: Membership of selected WGCNA modules was used for further analysis. (a-c) Three modules (turquoise, blue, and magenta) were selected from the GBM versus control analysis (G/C), and (d) one module (blue) was selected from the primary versus recurrent GBM analysis (R/PG). The figure shows the genes included in each of these selected modules.

Genes in the blue module of WGCNA G/C were enriched in “DNA binding” (MF), “DNA metabolic process” (BP), and “chromosome” (CC) (figure 9b). Similarly, genes in the magenta module of WGCNA G/C were mainly associated with “double-stranded RNA binding”,

“defense response to virus”, and “cytosol” in the (MF), (BP), and (CC), respectively (figure 9c). Enrichment analysis of genes of the blue module of the WGCNA R/PG indicated associations with “protein binding” (MF), “neurogenesis” (BP), and “cell periphery” (CC) (figure 9d).



Results of DEPs

Sample quality assessment using principal component analysis (PCA) demonstrated clear separation among groups (figure 11a). In addition, hierarchical clustering heatmap analysis revealed three distinct protein expression clusters (Figure 11b). Cluster1 consisted of proteins upregulated in the recurrence group but downregulated in the primary and control groups. Cluster 2 included proteins upregulated in the control group but downregulated in both primary and recurrent groups. Cluster 3 comprised proteins upregulated in the primary and control groups but downregulated in the recurrence group.

Based on protein-protein interaction network analysis, the top 10 hub proteins were identified for each cluster. In Cluster 1, *SNU13*, *RPS2*,

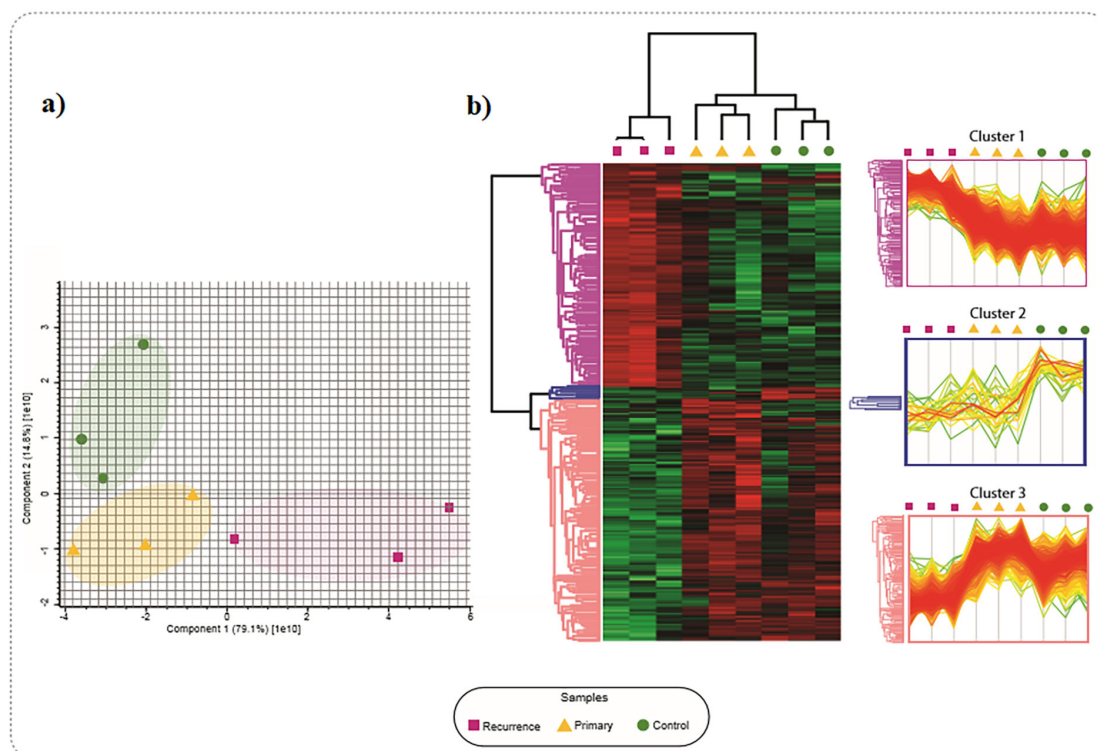


Figure 11: Proteomic analysis of GBM samples is illustrated. (a) The principal component analysis (PCA) used to assess the quality and distribution of recurrence, primary, and control samples; (b) a heatmap clustering proteins with similar expression patterns. Three clusters were identified based on protein expression patterns. Cluster 1 contained proteins that were upregulated in the recurrence group. Cluster 2 contained proteins that were upregulated in the control group. Cluster 3 contained proteins that were upregulated in both the primary and control groups.

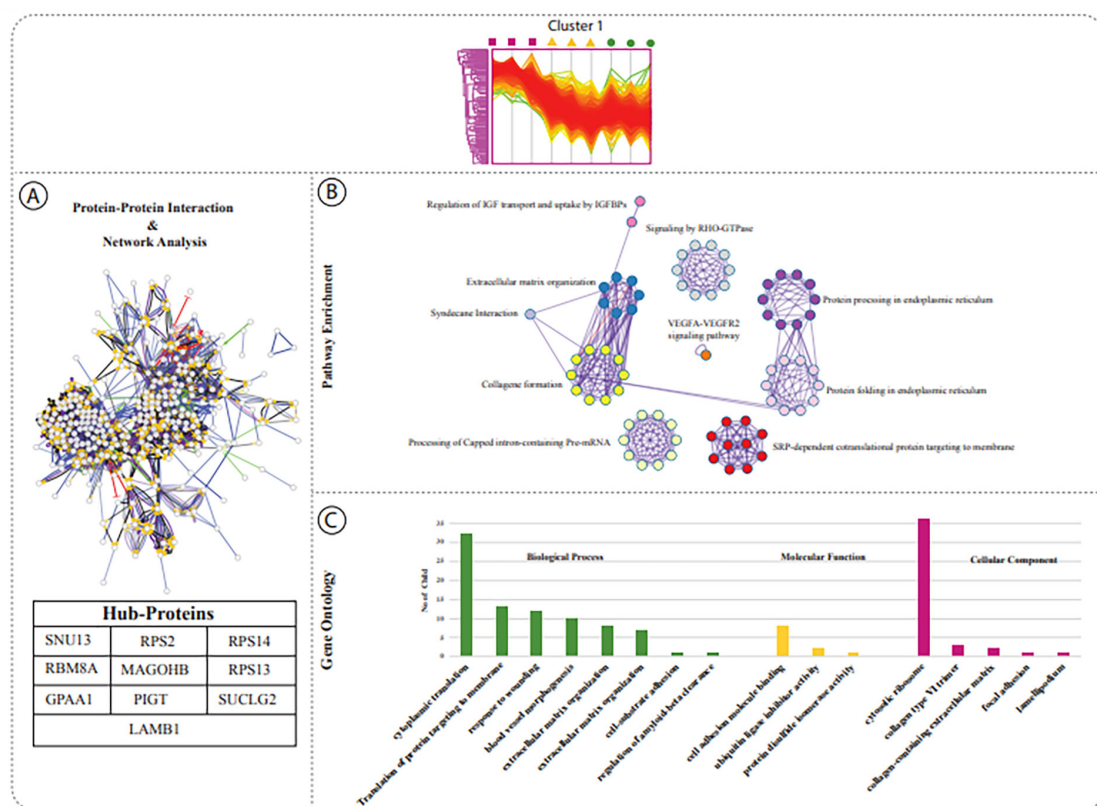


Figure 12: Enrichment analysis and protein-protein interaction of Cluster 1 are depicted. (a) The expression of the top 10 hub proteins identified from network analysis across recurrence, primary, and control groups; (b) the pathway enrichment of these proteins; and (c) the gene ontology analysis including biological process, molecular function, and cellular component for the hub proteins.

RPS14, *RBM8A*, *MAGOHB*, *RPS13*, *GPAA1*, *PIGT*, *SUCLG2*, and *LAMB1* proteins were increased in the recurrence group compared with the primary and control groups (figure 12a). Pathway enrichment analysis revealed that these proteins were mainly involved in “SRP-dependent co-translational protein targeting to membrane”, “Protein folding”, “Protein processing”, “Collagen formation”, “Processing of Capped intron-containing Pre-mRNA”, and “Signaling by RHO-GTPase” (figure 12b). Gene ontology (GO) analysis of the proteins of cluster 1 indicated enrichment in “cytoplasmic translation” (BP), “cell adhesion molecule binding” (MF), and “cytosolic ribosome” (CC) (figure 12c).

In Cluster 2, the 10 hub DEPs included *SKP1*, *FBXO2*, *FGG*, *RBCK1*, *PRDX6*, *UNC13D*, *CYB5R3*, *VAT1*, *MAP2K2*, and *HSPA2* were increased in the control group compared with primary and recurrent groups (figure 13a). Pathway analysis showed enrichment in “Glutathione conjugation”, “Negative Regulation of apoptotic Signaling”, “Regulation of Actin Cytoskeleton”, and “Signaling RHO GTPase” (figure 13b). GO analysis indicated associations with “Positive regulation of membrane

permeability” (BP) and “ADP binding”, “Oxidoreductase activity” (MF) (figure 13c).

In Cluster 3, *VAMP1*, *HSPA8*, *MAOK1*, *DNAJC6*, *DNM1*, *ITSN1*, *SYNJ1*, *AP2M1*, *DNM3*, and *SH3GL2* DEPs were increased in primary and control groups compared with the recurrence group (figure 14a). These proteins were enriched in pathways related to “Synaptic Vesicle Cycle”, “Carbon Metabolism”, “Membrane Trafficking”, “Retrograde endocannabinoid Signaling”, and “Opioid Signaling” (figure 14b). GO analysis demonstrated enrichment in “Vesicle-mediated transport” (BP), “Cytoskeletal protein binding” (MF), and related to “vesicle” (CC) (figure 14c).

Discussion

In the present study, we performed an integrative transcriptomic and proteomic analysis to identify key genes, proteins, and molecular pathways associated with GBM development and recurrence. By combining differential expression analysis, weighted gene co-expression network analysis (WGCNA), and proteomic profiling, we identified *ALDH2*, *ARRB1*, and *TPRG1L* as common genes associated

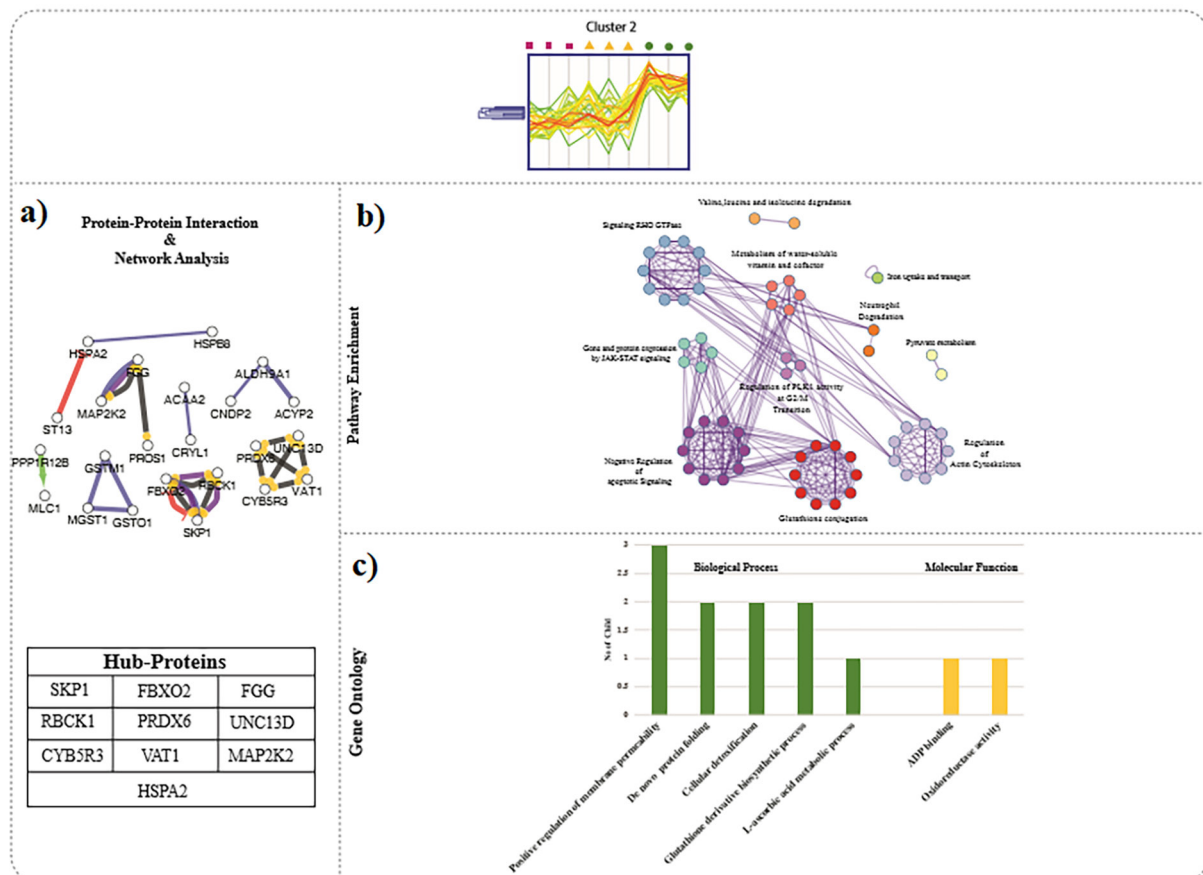


Figure 13: Enrichment analysis and protein–protein interaction of cluster 2 is presented. (a) The expression of the top 10 hub differentially expressed proteins identified by network analysis across recurrence, primary, and control groups; (b) The pathway enrichment analysis of these hub proteins; and (c) The gene ontology analysis, including biological process and molecular function, for the proteins in cluster 2.

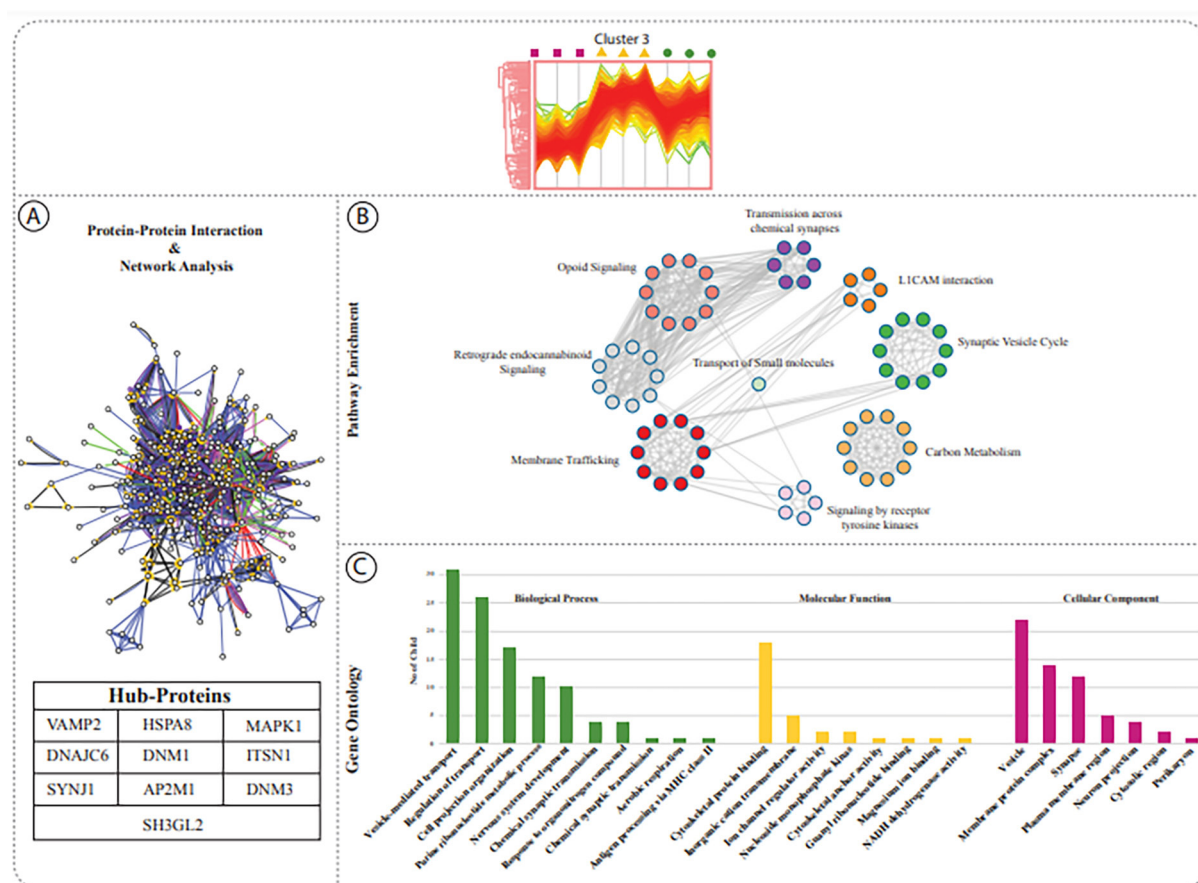


Figure 14: Enrichment analysis and protein–protein interaction of Cluster 3 are depicted. (a) The expression of the top 10 hub differentially expressed proteins identified by network analysis across recurrence, primary, and control groups; (b) The pathway enrichment analysis of these hub proteins; and (c) The gene ontology analysis, including biological process, molecular function, and cellular component, for the proteins in cluster 3.

with both GBM progression and recurrence. In addition, proteomic analysis revealed three distinct protein clusters and highlighted several biological processes and molecular functions potentially involved in GBM pathogenesis. Validation using an independent dataset confirmed the downregulation of *ARRB1* and *ALDH2* in GBM tissues compared with normal brain tissues. To our knowledge, few studies have integrated transcriptomic and proteomic data while simultaneously investigating genes associated with both GBM occurrence and recurrence. Therefore, our study helps address this knowledge gap and provides a comprehensive overview of the molecular alterations underlying GBM progression.

Because no publicly available dataset simultaneously included primary GBM, recurrent GBM, and control samples, we analyzed these groups separately. Nine datasets were used to investigate genes associated with GBM versus control samples, whereas one dataset was used to evaluate recurrent versus primary GBM. The integration of DEGs and WGCNA results enabled the identification of *ALDH2*, *ARRB1*,

and *TPRG1L* as key genes potentially involved in GBM biology. These findings suggest that *ALDH2* and *ARRB1*, in particular, may serve as promising biomarkers for GBM diagnosis and progression.

A previous study reported that ALDHs influenced the tumor cell cycle, epithelial-mesenchymal transition, and infiltration of immunocytes.²⁵ Moreover, our constructed prognostic model, a nomogram based on risk score and glioma type, could precisely predict overall survival outcome, indicating its potential clinical application value.

ALDHs are a family of 19 isozymes involved in oxidizing endogenous and exogenous aldehydes to carboxylic acids. *ALDHs* are dysregulated in diseases such as Alzheimer's disease, Parkinson's disease, and solid tumors.²⁶ Notably, in solid tumors (including GBM), *ALDHs* maintain a stem-cell-like phenotype, driving aggressive clinical progression. *ALDH3B1* and *ALDH16A1*, two main contributors to the scoring system, affect GBM cell proliferation and migration by inducing cell-cycle arrest and promoting epithelial-mesenchymal transition (EMT).²⁵

However, *ALDH1A1* is related to GBM proliferation and migration.²⁷ *ALDH2* plays a critical role in safeguarding cells against acetaldehyde toxicity and is closely linked to human metabolism.²⁸ *ALDH2* expression levels were analyzed in all 33 cancer types from the TCGA database and the Genotype-Tissue Expression (GTEx) database. The results showed that *ALDH2* expression was decreased in GBM.²⁹ Another study reported that the expression level of this gene was higher in low-grade glioma.³⁰

The association of the *ARRB1* gene with GBM was found in some studies, and it was reported that this gene in GBM has potential significance for therapeutic strategies targeting GBM through NK1R (Tachykinin receptor neurokinin-1).³¹ The expression level of *ARRB1* mRNA was lower in 17 tumor tissues, including GBM, compared to normal tissues according to datasets from the TCGA database.³² Another study also reported that both mRNA and protein levels of *ARRB1* were reduced in glioma cell lines. Immunohistochemistry data also demonstrated that *ARRB1* expression in glioma tissues was decreased compared to peritumoral brain tissue. Notably, *ARRB1* expression was more significantly reduced in stage III-IV gliomas than in stage I-II gliomas. Reduced *ARRB1* expression could serve as a biomarker to evaluate clinical and pathological features and predict prognosis in glioma patients.³³ However, the association of *TPRG1L* with GBM has not been investigated in previous studies.

Moreover, in the DEP analysis step, we found three clusters that were differentially expressed in the three groups (primary, recurrence, and control); then, 10 hub genes in each cluster were introduced.

Contrary to expectations, we did not find any common genes between the transcriptomics and proteomics analyses. This lack of concordance between mRNA and protein levels has been widely reported and can be explained by multiple biological and technical factors. mRNA expression does not necessarily correlate with protein abundance due to post-transcriptional regulation, differences in mRNA stability, translational efficiency, and protein degradation rates. In addition, post-translational modifications and protein turnover may alter protein abundance independently of mRNA levels.³⁴ Technical factors such as differences in detection sensitivity, dynamic range, and sample heterogeneity between RNA-seq and proteomics platforms may also contribute to this discrepancy. Therefore, the absence of shared genes between RNA and protein datasets reflects the complexity of gene

regulation in GBM. On the other hand, the gene ontology analysis of the hub proteins in each cluster and the genes of WGCNA modules showed that there are some common molecular functions and biological processes between the hub proteins and the genes of modules. For instance, "cell adhesion molecule binding" in the molecular function (MF) category of Cluster1 and "Cytoskeletal protein binding", "Inorganic cation transmembrane", and "Ion channel regulator activity" in the MF category of Cluster 3 are also enriched in the MF category of WGCNA R/PG (blue module) or in WGCNA G/C (turquoise modules). As well, "response to wounding", "blood vessel morphogenesis" and "cell-substrate adhesion" in the biological process (BP) category of Cluster 1, and "Regulation of transport", "Cell projection organization", "Nervous system development", "Response to organonitrogen compound", "Chemical synaptic transmission", and "Vesicle-mediated transport" in the BP category of Cluster 3 are also enriched in the BP category of WGCNA R/PG (blue module) or in WGCNA G/C (turquoise modules) biological process.

The stroma within solid tumors exhibits cellular phenotypes and signaling pathways associated with wound healing, leading to the intriguing concept that a tumor behaves like a wound that never heals. Some shared features between tumors and healing wounds include fibroblast recruitment, extracellular matrix (ECM) deposition, immune cell infiltration, neovascularization, and cellular plasticity.³⁵ On the other hand, although wound healing is essential for tissue repair, it can also contribute to GBM progression.

Blood vessel formation and remodeling are crucial for GBM growth and progression. Indeed, GBM is known for its high angiogenic activity. The extensive growth of blood vessels (angiogenesis) plays a crucial role in tumor progression and invasion. These newly formed blood vessels supply nutrients and oxygen to the rapidly dividing tumor cells, allowing them to thrive and infiltrate surrounding brain tissue. Understanding angiogenesis in GBM is essential for developing targeted therapies to disrupt this process and improve patient outcomes.³⁶

Altered cell-substrate interactions may impact GBM invasion and metastasis.³⁷ GBM often affects the nervous system, and aberrant neural development could impact prognosis.³⁸ However, GBM's impact on synaptic transmission could influence cognitive function and overall prognosis.³⁹

A limitation of this study is that all analyses were conducted using publicly available

datasets without independent experimental validation. Future studies are needed to confirm the functional roles of the identified genes and proteins and to explore their clinical utility.

Conclusion

ALDH2, *ARRB1*, and *TPRG1L* were identified as critical genes associated with GBM development and recurrence. In addition, several shared biological processes and molecular functions, including cell adhesion molecule binding, cytoskeletal protein binding, and ion channel regulator activity, were implicated in GBM progression. These findings provide new insights into the molecular mechanisms underlying GBM and may contribute to the identification of potential biomarkers and therapeutic targets. Further studies are required to validate these findings experimentally and to assess their clinical applicability in patients with GBM.

Acknowledgment

The authors acknowledge the Gene Expression Omnibus (GEO) and ProteomeXchange databases for providing the publicly available transcriptomic and proteomic datasets used in this study.

Authors' Contribution

N.K: Transcriptomic (DEGs and WGCNA) and proteomic analyses, data collection, and drafting the manuscript; A.T: Transcriptomic (DEGs and WGCNA) and proteomic analyses, data collection, and drafting the manuscript; H.H: Transcriptomic (DEGs and WGCNA) and proteomic analyses, data collection, and drafting the manuscript; M.V: Transcriptomic (DEGs and WGCNA) and proteomic analyses, data collection, and drafting the manuscript; H.Kh.M: Study design and reviewing the manuscript; M.Y: Conceptualization, study design, data analysis, data interpretation, and drafting the manuscript; All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Declaration of AI

No artificial intelligence (AI) tools were used in the design of the study, data analysis, interpretation of the results, or preparation of the manuscript.

Conflict of Interest: None declared.

References

- Ostrom QT, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014-2018. *Neuro Oncol.* 2021;23:1-105. doi: 10.1093/neuonc/noab200. PubMed PMID: 34608945; PubMed Central PMCID: PMC8491279.
- Wen PY, Weller M, Lee EQ, Alexander BM, Barnholtz-Sloan JS, Barthel FP, et al. Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro Oncol.* 2020;22:1073-113. doi: 10.1093/neuonc/noaa106. PubMed PMID: 32328653; PubMed Central PMCID: PMC7594557.
- Fekete B, Werlenius K, Tisell M, Pivodic A, Smits A, Jakola AS, et al. What predicts survival in glioblastoma? A population-based study of changes in clinical management and outcome. *Front Surg.* 2023;10:1249366. doi: 10.3389/fsurg.2023.1249366. PubMed PMID: 37711136; PubMed Central PMCID: PMC10498299.
- Prakash V, Gabrani R. Epigenetic dysregulation in glioblastoma: potential pathways to precision medicine. *Neurogenetics.* 2024;26:5. doi: 10.1007/s10048-024-00793-5. PubMed PMID: 39585441.
- Vaz-Salgado MA, Villamayor M, Albarran V, Alia V, Sotoca P, Chamorro J, et al. Recurrent Glioblastoma: A Review of the Treatment Options. *Cancers (Basel).* 2023;15. doi: 10.3390/cancers15174279. PubMed PMID: 37686553; PubMed Central PMCID: PMC10487236.
- Zhang T, Wang Y, Dong Y, Liu L, Han Y, Wang H, et al. Identification of Novel Diagnostic Biomarkers in Prostate Adenocarcinoma Based on the Stromal-Immune Score and Analysis of the WGCNA and ceRNA Network. *Dis Markers.* 2022;2022:1909196. doi: 10.1155/2022/1909196. PubMed PMID: 35075375; PubMed Central PMCID: PMC8783709.
- Chen L, Sun T, Li J, Zhao Y. Identification of hub genes and biological pathways in glioma via integrated bioinformatics analysis. *J Int Med Res.* 2022;50:3000605221103976. doi: 10.1177/03000605221103976. PubMed PMID: 35676807; PubMed Central PMCID: PMC9189557.

- 8 Perez-Riverol Y, Bai J, Bandla C, Garcia-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 2022;50:D543-D52. doi: 10.1093/nar/gkab1038. PubMed PMID: 34723319; PubMed Central PMCID: PMC8728295.
- 9 Jeanmougin M, Havik AB, Cekaite L, Brandal P, Sveen A, Meling TR, et al. Improved prognostication of glioblastoma beyond molecular subtyping by transcriptional profiling of the tumor microenvironment. *Mol Oncol.* 2020;14:1016-27. doi: 10.1002/1878-0261.12668. PubMed PMID: 32171051; PubMed Central PMCID: PMC7191188.
- 10 Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics.* 2010;26:2363-7. doi: 10.1093/bioinformatics/btq431. PubMed PMID: 20688976; PubMed Central PMCID: PMC2944196.
- 11 Oh YT, Cho HJ, Kim J, Lee JH, Rho K, Seo YJ, et al. Translational validation of personalized treatment strategy based on genetic characteristics of glioblastoma. *PLoS One.* 2014;9:e103327. doi: 10.1371/journal.pone.0103327. PubMed PMID: 25084005; PubMed Central PMCID: PMC4118874.
- 12 Sun L, Hui AM, Su Q, Vortmeyer A, Kotliarov Y, Pastorino S, et al. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell.* 2006;9:287-300. doi: 10.1016/j.ccr.2006.03.003. PubMed PMID: 16616334.
- 13 Gusev Y, Bhuvaneshwar K, Song L, Zenklusen JC, Fine H, Madhavan S. The REMBRANDT study, a large collection of genomic data from brain cancer patients. *Sci Data.* 2018;5:180158. doi: 10.1038/sdata.2018.158. PubMed PMID: 30106394; PubMed Central PMCID: PMC6091243.
- 14 Schuster A, Klein E, Neirinckx V, Knudsen AM, Fabian C, Hau AC, et al. AN1-type zinc finger protein 3 (ZFAND3) is a transcriptional regulator that drives Glioblastoma invasion. *Nat Commun.* 2020;11:6366. doi: 10.1038/s41467-020-20029-y. PubMed PMID: 33311477; PubMed Central PMCID: PMC7732990.
- 15 Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30:207-10. doi: 10.1093/nar/30.1.207. PubMed PMID: 11752295; PubMed Central PMCID: PMC99122.
- 16 Reinhardt J, Landsberg J, Schmid-Burgk JL, Ramis BB, Bald T, Glodde N, et al. MAPK Signaling and Inflammation Link Melanoma Phenotype Switching to Induction of CD73 during Immunotherapy. *Cancer Res.* 2017;77:4697-709. doi: 10.1158/0008-5472.CAN-17-0395. PubMed PMID: 28652246.
- 17 Kwon SM, Kang SH, Park CK, Jung S, Park ES, Lee JS, et al. Recurrent Glioblastomas Reveal Molecular Subtypes Associated with Mechanistic Implications of Drug-Resistance. *PLoS One.* 2015;10:e0140528. doi: 10.1371/journal.pone.0140528. PubMed PMID: 26466313; PubMed Central PMCID: PMC4605710.
- 18 Dunning M, Lynch A, Eldridge M. illuminaHumanv4.db: Illumina HumanHT12v4 annotation data (chip illuminaHumanv4). R package version. 2015;1.
- 19 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47. doi: 10.1093/nar/gkv007. PubMed PMID: 25605792; PubMed Central PMCID: PMC4402510.
- 20 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559. doi: 10.1186/1471-2105-9-559. PubMed PMID: 19114008; PubMed Central PMCID: PMC2631488.
- 21 Albert R. Scale-free networks in cell biology. *J Cell Sci.* 2005;118:4947-57. doi: 10.1242/jcs.02714. PubMed PMID: 16254242.
- 22 Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc.* 2016;11:2301-19. doi: 10.1038/nprot.2016.136. PubMed PMID: 27809316.
- 23 Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods.* 2016;13:731-40. doi: 10.1038/nmeth.3901. PubMed PMID: 27348712.
- 24 Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 2019;47:W191-W8. doi: 10.1093/nar/gkz369. PubMed PMID: 31066453; PubMed Central PMCID: PMC6602461.
- 25 Wang Z, Mo Y, Tan Y, Wen Z, Dai Z, Zhang H, et al. The ALDH Family Contributes to Immunocyte Infiltration, Proliferation and Epithelial-Mesenchymal Transformation in Glioma. *Front Immunol.* 2021;12:756606. doi: 10.3389/fimmu.2021.756606. PubMed

- PMID: 35116021; PubMed Central PMCID: PMC8805082.
- 26 Xia J, Li S, Liu S, Zhang L. Aldehyde dehydrogenase in solid tumors and other diseases: Potential biomarkers and therapeutic targets. *MedComm* (2020). 2023;4:e195. doi: 10.1002/mco2.195. PubMed PMID: 36694633; PubMed Central PMCID: PMC9842923.
- 27 Xu SL, Liu S, Cui W, Shi Y, Liu Q, Duan JJ, et al. Aldehyde dehydrogenase 1A1 circumscribes high invasive glioma cells and predicts poor prognosis. *Am J Cancer Res*. 2015;5:1471-83. PubMed PMID: 26101711; PubMed Central PMCID: PMC4473324.
- 28 Lamb RJ, Griffiths K, Lip GYH, Sorokin V, Frenneaux MP, Feelisch M, et al. ALDH2 polymorphism and myocardial infarction: From alcohol metabolism to redox regulation. *Pharmacol Ther*. 2024;259:108666. doi: 10.1016/j.pharmthera.2024.108666. PubMed PMID: 38763322.
- 29 Shen X, Yan Z, Huang Y, Zhu Q, Zhang G, Ci H, et al. ALDH2 as an immunological and prognostic biomarker: Insights from pan-cancer analysis. *Medicine (Baltimore)*. 2024;103:e37820. doi: 10.1097/MD.00000000000037820. PubMed PMID: 38640328; PubMed Central PMCID: PMC11030019.
- 30 Zhang S, Chen S, Wang Z, Li J, Yuan Y, Feng W, et al. Prognosis prediction and tumor immune microenvironment characterization based on tryptophan metabolism-related genes signature in brain glioma. *Front Pharmacol*. 2022;13:1061597. doi: 10.3389/fphar.2022.1061597. PubMed PMID: 36386216; PubMed Central PMCID: PMC9663932.
- 31 Zhang YX, Li XF, Yuan GQ, Hu H, Song XY, Li JY, et al. beta-Arrestin 1 has an essential role in neurokinin-1 receptor-mediated glioblastoma cell proliferation and G(2)/M phase transition. *J Biol Chem*. 2017;292:8933-47. doi: 10.1074/jbc.M116.770420. PubMed PMID: 28341744; PubMed Central PMCID: PMC5448126.
- 32 Ye Y, Jiang H, Wu Y, Wang G, Huang Y, Sun W, et al. Role of ARRB1 in prognosis and immunotherapy: A Pan-Cancer analysis. *Front Mol Biosci*. 2022;9:1001225. doi: 10.3389/fmolb.2022.1001225. PubMed PMID: 36213111; PubMed Central PMCID: PMC9538973.
- 33 Pu-shuai W. Expression and significance of arrestin beta 1 in patients with glioma. *Acta Anat Sin*. 2012;45:779-84.
- 34 Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012;13:227-32. doi: 10.1038/nrg3185. PubMed PMID: 22411467; PubMed Central PMCID: PMC3654667.
- 35 Foster DS, Jones RE, Ransom RC, Longaker MT, Norton JA. The evolving relationship of wound healing and tumor stroma. *JCI Insight*. 2018;3. doi: 10.1172/jci.insight.99911. PubMed PMID: 30232274; PubMed Central PMCID: PMC6237224.
- 36 Jhaveri N, Chen TC, Hofman FM. Tumor vasculature and glioma stem cells: Contributions to glioma progression. *Cancer Lett*. 2016;380:545-51. doi: 10.1016/j.canlet.2014.12.028. PubMed PMID: 25527451.
- 37 Perego C, Vanoni C, Massari S, Raimondi A, Pola S, Cattaneo MG, et al. Invasive behaviour of glioblastoma cell lines is associated with altered organisation of the cadherin-catenin adhesion system. *J Cell Sci*. 2002;115:3331-40. doi: 10.1242/jcs.115.16.3331. PubMed PMID: 12140264.
- 38 Zhang Z, Lin CC. Taking advantage of neural development to treat glioblastoma. *Eur J Neurosci*. 2014;40:2859-66. doi: 10.1111/ejn.12655. PubMed PMID: 24964151.
- 39 Venkataramani V, Tanev DI, Kuner T, Wick W, Winkler F. Synaptic input to brain tumors: clinical implications. *Neuro Oncol*. 2021;23:23-33. doi: 10.1093/neuonc/noaa158. PubMed PMID: 32623467; PubMed Central PMCID: PMC7850064.